

Weekly Report

July 1st, 2018

Done:

1. Since Prof. Ma is busy in Taiwan and didn't reply to my email about the data processing organization/storytelling topic and design, most time of this week was spent on coding.
2. Discuss on one of the fire data source. Some data are not updated timely now, so we have to find alternatives, so we explored some data on NASA website.

To Do:

1. Try to apply some methodologies (see paper reading 3) to enhance some designs of the data processing organization. Try to grab one guy for more discussion if Prof. Ma is not back.
2. Some modules (written by others) have been modified, I will update them in my system correspondingly.

Paper reading:

1 KDD2018 *Exact and Consistent Interpretation for Piecewise Linear Neural Networks: A Closed Form Solution* 本文对于分段线性神经网络做解读. 做这个类型神经网络的解读相对简单, 因为其激活函数都是像 ReLU 一样的分段线性函数. 该方法将模型的参数作为特征来寻找共享参数的输入输出数据(他们会被包在一个高维的多面体中), 以此来对模型进行解读. 相比我们之前熟知的 LIME 而言, 精确性有了保证; 另外据说分段线性函数也是可以逼近其他非线性激活函数的, 但是以我现在所了解的数学知识而言推广出去有点难度.

2 EuroVis2018 *Interactive Analysis of Word Vector Embeddings* 如果光从可视设计的角度而言, 设计是比较简单的, 但是文章对于 word vector embedding 中实际用到任务的调研归纳最后提炼到核心的几点是十分详尽的(而且主要是基于文献阅读而不是咨询专家), 而其设计正是基于这些任务, 设计了一些小而精的组件. 这些任务很多在 coursera 的在线课程中也有提到, 我也有所体会; 据此设计的小组件, 虽然都不是一个联动的可视化系统但是实用.

3 CHI2018 *Design Patterns for Data Comics* 是一篇综述也是一篇方法论, 对 data comics 设计有指导意义. 提到各种 design pattern 和 design space, 甚至可以是作为一种自动化生成信息图所套用的模板(可以再联系 Jeff Heer 那篇关于 narrative visualization 的综述)

4 CVPR2018 *Referring Relationships* 这篇文章利用 subject-predicate-object 这样的句式来对图像中的东西进行消歧义. 利用 predict 进行 Attention shift, 从图中一个东西转移 attention 而发现另一个东西, 当然也可以依此从另一个东西反推验证(通过不断迭代, 训练出这样一个模型). 这种方法除了消歧, 还能顺带根据这种 shift 来找出一些不容易识别的东西的位置. 有一个拍脑袋的想法是, 如果用户要拿可视分析系统做推理和决策, 他也会有类似的思考过程(SVO 句式), 若能代替用户做一些这样的事情, 可以一定程度上减少他们的认知负担; 甚至说不定也能像文章一样找到那些不仔细看发现不了的模式.

5 KDD2018 *NGUARD: A Game Bot Detection Framework for NetEase MMORPGs* 网易在外挂检测中用到的方法最终形成了一套框架, 并用于多个不同类游戏中, 取得了很好的效果. 其方法中结合了大量 RNN 方法, 读来有点吃力. 在检测变种外挂的实践中, 他们还是采用了聚类的方法, 结合前面 RNN 离线训练的 embedding 来探测变种外挂, 这种数据挖掘方法可以借鉴.